Did you see what I saw?: Comparing attentional synchrony during 360° video viewing in head mounted display and tablets

Farmer, H[1,2,3*], Bevan, C.[4], Green, D.[5], Rose, M.[5], Cater, K.[4], Stanton Fraser, D.[1]

[1] University of Bath, Bath, UK

[2] University of Greenwich, London, UK

[3] University College London, London, UK

[4] University of Bristol, Bristol, UK

[5] University of the West of England, Bristol, UK

*Corresponding Author: h.farmer@gre.ac.uk, School of Human Sciences, University of Greenwich, London, SE10 9LS, UK

Word Count: 6711

Acknowledgment: This work was supported by the EPSRC Virtual Realities - Immersive Documentary Encounters project - EP/P025595/1

Abstract

Advances in head mounted displays (HMDs) have increased the interest in cinematic virtual reality as an artform. However, the freedom of a viewer in 360 video presents challenges in ensuring that audiences do not inadvertently miss important events and locations. We examined whether the high level of immersion provided by HMDs encourages participants to synchronise their attention during viewing. Sixty-four participants watched the 360° documentary "Clouds Over Sidra" using either an HMD or via a flat screen tablet display. We used inter-subject correlation (ISC) analysis to measure attentional synchrony over the course of the video and to examine whether spatial and temporal factors led to different amounts of correlation both within and between groups. We found significantly greater ISC for the HMD compared to the tablet group. This effect was greatest for scenes with a unidirectional focus and at the start of scenes. We discuss our results in terms of the visual properties and the motor affordances of HMDs vs tablets. Our results show the value of HMDs in increasing attentional synchrony and may provide producers of 360° content insight in how to encourage or discourage synchronisation of viewing direction.

Public Significance Statement

The rapid uptake of VR by producers of nonfiction reflects a fascination with the potential of 360 media for viewer engagement. While VR allows viewers a novel freedom to find their own visual pathway through a given scene, at the same time producers have specific information that they need to convey. The power to elicit synchronization of viewers' attention is often held up as a key indicator of a content creator's ability to lead an audience through a narrative. In this paper we have shown that 360° video can elicit a degree of synchronisation between viewers and that this effect is increased when the video is viewed via a HMD when compared to being viewed on a flat screen tablet.

Introduction

The rapid advances in virtual reality (VR) technology in the last few years has led to increased interest in VR's potential as a means of delivering narrative media. Since late 2015, the on-going release of affordable, high specification head mounted display (HMD) systems including Oculus's Rift (2016) and Quest (2019), Samsung's GearVR (2015) and HTC's Vive (2016) has driven a vast increase in VR content. Release of English language nonfiction VR pieces alone increased from under 40 pieces in 2015 to over 180 in 2017 (Bevan & Green, 2017), and recent projections suggest that the VR and AR market will be worth $170 billion by 2022 (consultancy.uk, 2018).

Despite this enthusiastic uptake, the unique nature of VR presents significant artistic and technical challenges for content production and delivery. Within nonfiction production, 360 video is widely used. In this content production model, footage from a series of cameras set in a circular array are stitched together to form one panoramic scene. This scene is then projected onto a spherical surface and then usually presented via an HMD where directional data is used to determine what section of the video ('viewport') to display. Producers transitioning from traditional filmmaking in 2D into the world of virtual narrative have often been drawn to the use of 360° video due to similar skill requirements to traditional film making, high visual fidelity, platform adaptability and relatively low cost. A recent analysis of nonfiction VR (Bevan et al., 2019) suggests that around three quarters of all English language nonfiction VR content made between 2012 and 2018 were either solely or predominately composed of 360° video.

However, the omnidirectional nature of 360° video inevitably reduces the amount of control media producers have over the direction of viewers' attention, leading for calls for the

development of "a new screen grammar" (Dooley, 2017) and greater understanding of "the geometry of story-telling" (Pope et al., 2017). In addition to these issues around content production, greater knowledge of where viewers are likely to direct their gaze could also help to optimise the storage, transmission and rendering of 360° videos (Ozcinar & Smolic, 2018). In this context, an increased understanding of how people visually navigate through 360° videos has the potential to yield important insights for both content producers and tech developers working in VR.

One source of insight in designing compelling VR scenes is to examine previous research on viewing attention in the perception of dynamic scenes. Research in this area has used eye tracking technology to identify the factors that contribute to gaze fixation, taken as a proxy for attention. Using this approach, researchers have demonstrated that a variety of factors can modulate the pattern of viewer gaze. These include low level salience (Carmi & Itti, 2006; Itti, 2005), the motion of objects or people within the scene (Mital et al., 2011), task requirements (Hutson et al., 2017) amount of social content (Birmingham et al., 2009; Coutrot & Guyader, 2014; Rubo & Gamer, 2018), temporal order (Wang et al., 2012) and emotional valance (Rubo & Gamer, 2018; Subramanian et al., 2014).

Perhaps the most effective measure for assessing the success of content producers in driving viewing behaviour in 360° video is the extent to which different viewers attend to the same area of a scene. Such attentional synchrony (AS) plays a key role in successful story telling by allowing directors and editors to ensure that viewers are attending to the part of the visual scene that is most critical to the narrative at each particular time (Smith, 2013; Smith et al., 2012). Previous research has shown the role of several factors in modulating AS when watching traditional 2D films. Mital et al. (2011) showed that motion was a key factor in determining how much different individuals gazes clustered around similar points, a finding

supported by the fact that AS is higher when viewing dynamic as opposed to static scenes (Smith & Mital, 2013). AS has also been shown to be greater for free as opposed to task driven viewing (Smith & Mital, 2013) and to decrease during repeated viewing of a scene (Breathnach, 2016) suggesting that AS may be driven by the spontaneous extraction of novel scene features. Other studies have demonstrated that top down factors like viewer context (Loschky et al., 2015) can also influence AS, and AS has also been shown to have developmental and evolutionary components with higher synchrony in adults compared to children (Franchak et al., 2016; Kirkorian & Anderson, 2018) and in humans compared to monkeys (Shepherd et al., 2010).

One consistent and particularly relevant finding is that AS is stronger for tightly cut videos than for naturalistic scenes (Dorr et al., 2010; Hasson, Furman, et al., 2008). Temporal order also affects AS with greater AS when viewing structured compared to randomly edited sequences (Kirkorian & Anderson, 2018). These findings demonstrate the importance of techniques such as continuity editing, and close ups in directing the viewers gaze towards salient narrative events. The lack of many of these techniques in 360° video presents a challenge for content producers in ensuring that audience members perceive the story as intended, rather than fixating on areas of the scene without meaningful content.

A variety of approaches have been used to measure AS including: (a) bivariate contour ellipse area (Kirkorian & Anderson, 2018), which assesses the size of ellipse needed to capture all participants' gaze coordinates; (b) normalized scan path saliency (Dorr et al., 2010), which measures the correspondence between saliency maps and ground truth, computed as the average normalized saliency at fixated locations; (c) Gaussian mixture modelling (Mital et al., 2011; Smith & Mital, 2013), which probabilistically represents the clustering of eye movements across subjects; and (d) inter-subject correlation (ISC;

Burleson-Lesser et al., 2017; Franchak et al., 2016; Shepherd et al., 2010), in which each participant's motion path is separately correlated with that of every other participant before being averaged together to create a mean ISC for that participant. In the current study we chose to use ISC as this approach allows for the calculation of each individual's average AS with all other group members. This enables us to examine individual difference in AS while the other approaches can only characterise AS at the whole group level (Franchak et al., 2016).

ISC has been previously used to examine synchrony in neural responses to naturalistic film (Adolphs et al., 2016; Hasson et al., 2004; Hasson, Landesman, et al., 2008; Herbec et al., 2015). Hasson and colleagues (2008) employed functional magnetic resonance imaging (fMRI) and eye tracking and found that viewing tightly edited films such as Sergio Leone's "The Good, The Bad, and The Ugly**"** led to greater ISC for both viewer's eye movements and their neural response in brain areas associated with vision and attention than unedited footage. This suggests that ISC is a strong marker of how much control a director has over audience attention. Increased ISC in neural responses has been linked to improved memory encoding of events (Hasson, Furman, et al., 2008) and the amount of ISC in small groups of viewers has been shown to reliably predict the preferences of thousands (Dmochowski et al., 2014). Further work has expanded the use of ISC outside of fMRI data to show synchronisation of response in EEG (Poulsen et al., 2017), MEG (Lankinen et al., 2014) and physiological (Bracken et al., 2014; Golland et al., 2014) measures. ISC of eye movements has also been found to correlate with gaze salience (Franchak et al., 2016) indicating that it may act as a reliable indicator of visual salience.

To date only two studies (Bender, 2018; Sitzmann et al., 2018) have examined viewer synchrony in 360° video. Bender (2018) carried out a qualitative assessment of heat maps

based on the head direction of participants who watched a narrative 360° video and used a qualitative approach to classify the maps according to the concentration of attention within the scene. Bender found that most scenes displayed strong AS which appeared to be driven by the salient events such as a character speaking. Sitzmann and colleagues (2018) took a more quantitative approach tracking the head and eye movements of participants wearing an HMD while viewing 22 static 360° scenes. They calculated the receiver operating characteristic curve of each participant's fixations compared to the fixations of all other participants and found a high level of AS comparable to previous findings from 2D scenes. However, both studies had limitations in terms of measuring AS to dynamic scenes. Bender's (2018) analysis was purely qualitative and did not statistically test the amount of synchrony across the entire video, while the use of static scenes by Sitzmann and colleagues (2018) limits how far their results can be extrapolated to the dynamic scenes found in most 360° videos.

One factor that might alter the amount of AS when watching 360° video is the device on which the video is displayed. While the development of commercial HMDs has been a key driver of 360° video content production, one of the appeals of 360° video compared to other forms of virtual reality is that it can easily be adapted for non-HMD formats including phones and tablets, thus broadening the audience for 360° video beyond those with access to HMDs (Zoric et al., 2013).

Three previous studies have contrasted viewing of 360° videos on 2D displays and HMDs. Sitzmann and colleagues (2018) found no difference in the similarity of fixation locations when participants watched still 360° scenes using a 2D desktop display when compared to using HMD. The other two studies did not collect gaze data but did examine viewer responses to different display conditions. MacQuarrie and Steed (2017) compared viewing a

full 360° video in an HMD with viewing a non-360° version of the video on either a normal desktop display or a display-plus-peripheral projections. The authors used a variety of measures including whether attention could be more easily guided towards important items in the 2D screen conditions compared to HMD but found no significant difference in their measure of subsequent recall of cued events suggesting an equal amount of attentional control between the conditions. Finally Passmore and colleagues (2016) used qualitative interviews to assess differences between watching 360° video in an HMD or on a smartphone, or watching a 2D version on a desktop screen. They then coded these interviews and found that overall attention was high in all three conditions. However, these studies all had limitations that narrow the conclusions that can be drawn from them. Both MacQuarrie and Steed and Passmore et al. relied on user reports rather than directly comparing AS, while Sitzmann and colleagues did not examine dynamic scenes in which HMD users might be more influenced by motion cues. It is therefore unclear whether AS in 360° video differs between display devices.

The current study examined whether the type of device used to deliver the video and specific features within the 360° video affect synchronisation between viewers. Due to the difficulties in deploying eye-tracking devices within an HMD, and given the finding of Sitzmann and colleagues (2018) of a strong coupling between gaze and head movements when viewing 360° scenes, we did not directly track gaze but rather used directional data from the HMD/Tablet with the central point in the scene as a proxy for gaze. We calculated the within group ISC for participants' movements in the pitch and yaw axes as they watched a 360° video using either an HMD or a tablet in motion-tracked 'magic window' mode. We also investigated how both the spatial and temporal aspects of viewed scenes affected ISC and how this interacted with the type of device used.

Method

Participants

Sixty-four participants (mean age ± SD: 25.82 ± 11.31 years; 25 males) gave their written informed consent to participate and were paid for their participation. Thirty-two participants were assigned to the HMD and Tablet groups respectively. Due to the lack of previous research examining ISC between different devices, sample size was calculated using G*Power (Faul et al., 2007) based on having the power to detect a small (d = 0.25) interaction between viewing device and the temporal/spatial quality of scenes. All participants were screened for previous history of epilepsy or brain injury and for family history of epilepsy prior to taking part in the study. All participants had normal or corrected to normal vision without the need for glasses. Ethical approval for the study was given by the University of Bath's Department of Psychology Research Ethics Committee.

Design

The experiment had a between subjects design with one independent variable which was the device used to watch the 360° documentary (HMD vs Tablet). During the study we measured pitch and yaw rotational data, captured from the on-board orientation sensors of the HMD and Tablet computer and used those values to derive the great circle distance of each coordinate pair from a reference point of [0°,0°].

Materials

A tablet rather than a mobile phone or cursor controlled desktop setup was used as the control condition as this allowed for the presentation of 360 video in a flat screen format that most

closely matched the affordances of the HMD (i.e. requiring whole body movement to rotate in yaw and with mechanistic restrictions on the amount of possible rotation in pitch rather than simply depending on cursor movement).

In order to habituate participants to the affordances of the viewing device they were using, all participants first viewed an introductory 2:20 minute cut from the 360° video 'Nature Meditation' (Eco VR, 2017) which shows scenes of nature. For the main task participants watched the 8:35 minute 360° documentary 'Clouds Over Sidra' (VRSE.works, 2015). 'Clouds over Sidra' provides a tour of a refugee camp with Sidra, a young girl from Syria, as a guide. Sidra shows her family, her makeshift classroom, as well as other parts of the camp, and talks about her life in the camp and hopes for the future. The documentary offers 360-degree immersion in the settings featured in the documentary and directed sound corresponding to the view. Participants in the HMD condition viewed a higher resolution (3840px x 1920px) version of the film, presented in the monoscopic equirectangular projection format with no stereoscopic separation. Due to the limits of available screen resolution, participants in the Tablet condition viewed a slightly lower resolution (2732px x 1366px) version of the film. However, as the HMD splits and duplicates visual content to present a full image to each eye separately the perceived image quality of the film was – in practice - roughly equivalent across the HMD and tablet conditions.

The hardware setup for the HMD condition consisted of one laptop PC station (Intel Core i7-7700HQ 2.80 GHz CPU, 16 GB of RAM, Nvidia GeForce GTX 1070 Graphics card). The HMD was an Oculus Rift (consumer model), with a field of view of 110° nominal, a resolution of 1080×1200 pixels per eye, and a refresh rate of 90Hz. In the tablet condition participants viewed the documentary on a Samsung Galaxy Tab S 10.5 (Model SM-T80016, Exynos 5 Octa 5420 1.9 GHz CPU, 3GB of RAM, Mali-T628 MP6 Graphics). This tablet has

a screen resolution of 2560px x1600px, a field of view of approximately 85° and a refresh rate of 60Hz. Audio in the Tablet condition was delivered via Sony MDR-ZX330BT Bluetooth wireless headphones.

In both setups, participants viewed the 360° videos using a custom-built 360° video player application that was created by the researchers using the Unity software development package (version 2017.3.1f1). Unity is a cross-platform development environment, allowing the same codebase to be shared between the HMD and the Tablet. Playback of 360° video content was facilitated using the in-built video player component provided by Unity, and was rendered using the Skybox Panoramic Shader method detailed in Margerie (2018).

In addition to our head tracking measures we also measured engagement with the documentary using a seven item scale that Schutte and Stilinović (2017) adapted from Wiebe, Lamb, Hardy, & Sharek (2014) which included statements such as “The time just slipped away” and “I lost track of the world around me”. Responses were given on a Likert scale from 1 to 5 and the response across questions were averaged together to create a mean engagement score.

Procedure

Participants were randomly assigned to either the HMD or Tablet condition and verbally briefed about the study. Participants in the HMD condition were given additional information about how to adjust the headset straps and inter-ocular distance for viewing comfort while those in the Tablet condition were shown how the motion-tracked “magic window” set up of the tablet worked. All participants viewed the 360° videos while seated on a swivel chair that allowed easy rotation in the yaw dimension. For participants in the HMD condition the viewpoint in the yaw dimension could be changed by moving either their head or their whole

body to either side. The viewpoint in the pitch dimension could be changed by tilting their head up or down. For participants in the tablet condition the viewpoint in the yaw dimension could be changed by rotating either the tablet itself, or their whole body while holding the tablet, from side to side in the horizontal plane. The viewpoint in the pitch dimension could be changed by rotating the tablet, held in their hands, up or down.

The study was conducted in the CREATE lab at the University of Bath, a large room which contained a number of desks containing computers, a small sink area in the corner nearest participants and a set of floor to ceiling windows offering a view of the Bath campus. While using the device participants were concealed from the experimenter by a screen so that they did not feel too self-conscious to visually explore the environment. After watching the short introductory video, direction tracking was turned on and participants watched "Clouds Over Sidra". Having watched the piece participants completed a number of questionnaires including the measure of engagement. Finally, participants were thoroughly debriefed and given a detailed sheet explaining the purpose of the experiment.

Data Analysis

Pre-processing

For every rendered frame of playback, the current pitch and yaw orientation of the main scene camera was captured as a Euler angle and logged in a text file. This resulted in a sampling rate of 90Hz for the HMD, and a slightly reduced rate of 60Hz for the less powerful Tablet computer. Each frame was further labelled with the current time elapsed (msec) and the frame number. Orientation values were captured from the viewport camera using the transform class provided by Unity (Camera.main.transform.eulerAngles.x and Camera.main.transform.eulerAngles.y).

For the tablet condition, the 360° video was rendered to a full screen monoscopic viewport, with the orientation of the virtual camera synchronised to the device's internal gyroscope. This method of viewing 360° video is commonly referred to as "Magic Window", and provides an alternative method of viewing 360° video content (while retaining the freedom to move the viewport) in the absence of a Virtual Reality HMD.

By default, the inbuilt gyroscope of the tablet is oriented to assume that the tablet is placed flat on a table. To place the device in "magic window" mode, the baseline pitch of the camera required adjustment prior to playback to allow the participant to hold the device at eye level. To achieve this, participants were provided with a button to calibrate the viewport of the tablet by adjusting the pitch of the camera in 5-degree increments until an artificial horizon was centred in the middle of the screen. This adjustment was then stored and subsequently applied as an offset to each playback frame. For the HMD, no calibration was required, and no additional pre-processing was applied to the orientation data.

In both the HMD and tablet versions, the sampling rate of the viewport comfortably exceeded the frame rate of the video. Prior to analysis, tracking data was down-sampled to a rate of 10Hz to remove redundant and duplicated data points. Since our main interest was how participants synchronised their head movements while watching naturalistic scenes, we then further cut the time series to remove the opening and closing credits, leaving a time series with a total time of 7:16.2 minutes, starting at 9.1 seconds into the original video and ending at 7:25.5 minutes into the video.

Descriptive Statistics

Due to the directional nature of our statistics we present our descriptive statistical data in the format recommended by Cremers and Klugkist (2018) using the R package "circular"

(Agostinelli & Lund, 2017). We give values in each condition and dimension for four variables. First, the mean direction ($\bar{\theta}$) which is found by converting each circular datapoint into a vector composed of the sine and cos of that datapoints value in degrees, adding all the resulting vectors head to toe, and finding the direction of a new vector connecting the toe of the first vector to the head of the last vector. Second, the mean resultant length ($\bar{R}$) which is found by taking the length of the new vector described above, meaning that a lower value represents greater dispersion in datapoints. Third, the circular variance (Vm) which is the opposite of the mean resultant length and finally the circular standard deviation (v) where higher values indicate greater dispersion. To test differences between the distribution of mean pitch and yaw values between our device groups we report the outcomes of Watson-Wheeler tests.

We used the pitch and yaw data outputted from our player to calculate the great-circle distance (GCD) between each coordinate pair and the starting point of [0°,0°] using the Haversine method. This represents the shortest distance between two points on the surface of a sphere, measured along the surface (as opposed to a straight line through the sphere's interior, see Figure 1). To do this we first converted our yaw data into longitude by re-centring it so that values above 180° had 360° subtracted from them (meaning that the range of value now went from -180° to 180° with 0 retaining its original value. Next the pitch data was converted into latitude by making pitch values below 90° negative, subtracting 180° from values between 90° and 270° and subtracting 360° from values above 270° and then making that value positive. Following this we used the distHaversine function from the R package "geosphere" (Hijmans et al., 2019) to calculate GCD of each pair of coordinates in each participants time series relative to [0°,0°] using an arbitrarily set radius value of one. This approach allowed us to focus on the amount of ISC between participants from a set reference point rather than the relative change in direction at each point of their time series.

We report the results of independent sample t-tests and Levene's test for homogeneity of variance which examined differences between the mean GCD of our two device groups.

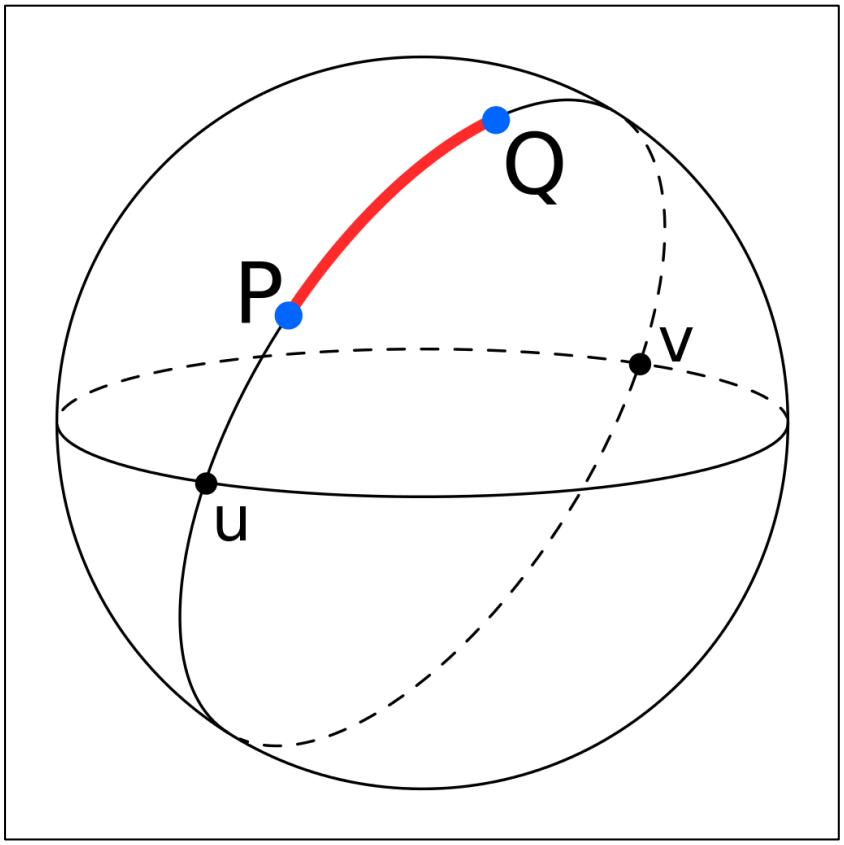

Figure 1. A diagram illustrating great-circle distance (drawn in red) between two points on a sphere, P and Q. Two antipodal points, u and v, are also shown (reprinted from Wikipedia (2020)) .

Analysis of Overall ISC

In order to calculate ISC, we first split our participants between the HMD and Tablet group. Next, we calculated the Pearson product moment correlation coefficient (ρ) of each participant's GCD data with the GCD data of every other participant in their device condition. To avoid underestimation due to the skewed sampling distribution of the correlation coefficient (Silver & Dunlap, 1987) we then converted these correlations into Fisher-transformed, z-normalised coefficients using the following formula:

$$z' = 0.5\log\left(\frac{1+\rho}{1-\rho}\right)$$

The z-normalised coefficients of each participant's data with that of all other participants were then averaged and the resulting mean z-transformed coefficients were transformed back into ρ to gain the mean correlation coefficient of each participant's tracking data with all other participants in their group, i.e. their ISC. We then compared whether the ISC for each group differed from 0 using one sampled t-tests and whether ISC differed between the groups using Welch's Two Sample t-tests. For descriptive purposes we also carried out an analysis of ISC for the whole period split across 10 second bins (see Figure 4B).

Analysis of Each Scene's ISC

In order to examine whether different types of scene showed greater correlation between participants in either group or between participant groups, we split our GCD data into time series based on the start and end point of each scene and then calculated each participant's ISC relative to their group. We then qualitatively categorised the scenes according to how far the scene's features directed the viewer's attention to a central point. Three categories were used: 1) Unidirectional: scenes in which there is one central focus that could be viewed within one 110° window. In all cases this was either an individual person or group of people; 2) Intermediate: scenes in which there was more than one location containing people or objects of interest, but where at least 110° of the frame contained no objects of interest. 3) Panoramic: Scenes in which there were people or objects of interest spread across the full 360° panorama, or where there were no clear objects of interest. Figure 2 shows two examples of each of the scene categories, while Table 1 gives a brief description of each scene, the category in which it was placed and its start and end times. To test whether the amount of ISC differed across these different scene types (and if the type of device used

influenced that difference), we first calculated each participant's mean fisher-transformed, z-scored ISC for each scene using the method described above. We then averaged the scenes in each category and transformed the z-scored means back into Pearson's ρ. We then ran an ANOVA comparing levels of ISC for each scene type.

Analysis of ISC Within Scenes

We were also interested in examining how ISC changed according to time period within each scene. To investigate this, we first calculated each participant's mean fisher-transformed, z-scored GCD ISC over 10 seconds after the beginning, around the mid-point and before the end of each scene. We then averaged together the ISCs across scenes for each time period and then transformed the z-scored means back into Pearson's ρ. We then ran an ANOVA comparing levels of ISC for each time period.

Analysis of Relationship between ISC and Engagement

To examine the effect of Device on engagement we ran a Welch's Two Sample t-test. To examine the relationship between ISC and engagement we carried out a linear regression with GCD ISC as the respective predictor and mean engagement score as the outcome variable.

Table 1. Scene numbers and description along with categories used in the scene analysis and start and end times used to define each scene time series.

Scene	Description	Category	Start (m:s)	End (m:s)	Length (m:s)
1	View of desert	Panoramic	0:09.1	0:33	00:23.9
2	Sidra in bedroom	Unidirectional	0:33.1	0:50.9	00:17.8
3	Sidra's family in house	Bidirectional	0:51	1:10	00:19.2
4	Journey to school	Panoramic	1:10.3	1:38.8	00:28.5
5	Classroom	Intermediate	1:38.9	2:08.6	00:29.7
6	Bakery	Intermediate	2:08.9	2:44	00:35.2
7	Children not in school	Unidirectional	2:44.1	3:04.8	00:20.7
8	Computer Room	Panoramic	3:04.9	3:48	00:43.1
9	Gym	Panoramic	3:48.1	4:13.5	00:25.4
10	Wrestling	Intermediate	4:13.6	4:44.3	00:30.7
11	Football Pitch	Intermediate	4:44.4	5:33.1	00:48.7
12	Family dinner	Unidirectional	5:33.2	6:16.1	00:42.9
13	Crowd of children	Panoramic	6:16.2	6:41	00:24.8
14	Sidra's bedroom	Unidirectional	6:41.1	6:55.6	00:14.5
15	View of Camp	Panoramic	6:55.7	7:25.3	07:25.3

Figure 2. Examples of unwrapped 360° video from each of the three categories used in the scene ISC analysis.

Results

Descriptive Statistics

Table 2 shows the descriptive statistics for each group. Due to the directional nature of our statistics we present our data in the format recommended by Cremers and Klugkist (2018). As can be seen the mean directions ($\bar{\theta}$) for both pitch and yaw were further from 0 in the Tablet group compared to the HMD group, suggesting that the Tablet group deviated further from the starting position than the HMD group (see Table 1 and Figure 3). In addition, the Tablet group show a smaller mean resultant length ($\bar{R}$) for both yaw and pitch suggesting greater spread in the mean direction in the Tablet group than in the HMD group. Both groups also showed a smaller $\bar{R}$ on the yaw compared to pitch axis.

To further investigate these differences, we carried out separate Watson-Wheeler tests on the data from each axis in order to assess whether the distribution of $\bar{\theta}$ between the device groups indicated that these group could be considered to be drawn from the same population. The

test for the pitch axis showed a significant difference in $\bar{\theta}$ distribution between the HMD and Tablet groups $W(2) = 35.83$, $p < .001$, $\eta^2 = 0.42$. The test for the yaw axis also showed a significant difference in $\bar{\theta}$ distribution between the HMD and Tablet groups $W(2) = 6.37$, $p = .041$, $\eta^2 = 0.15$.

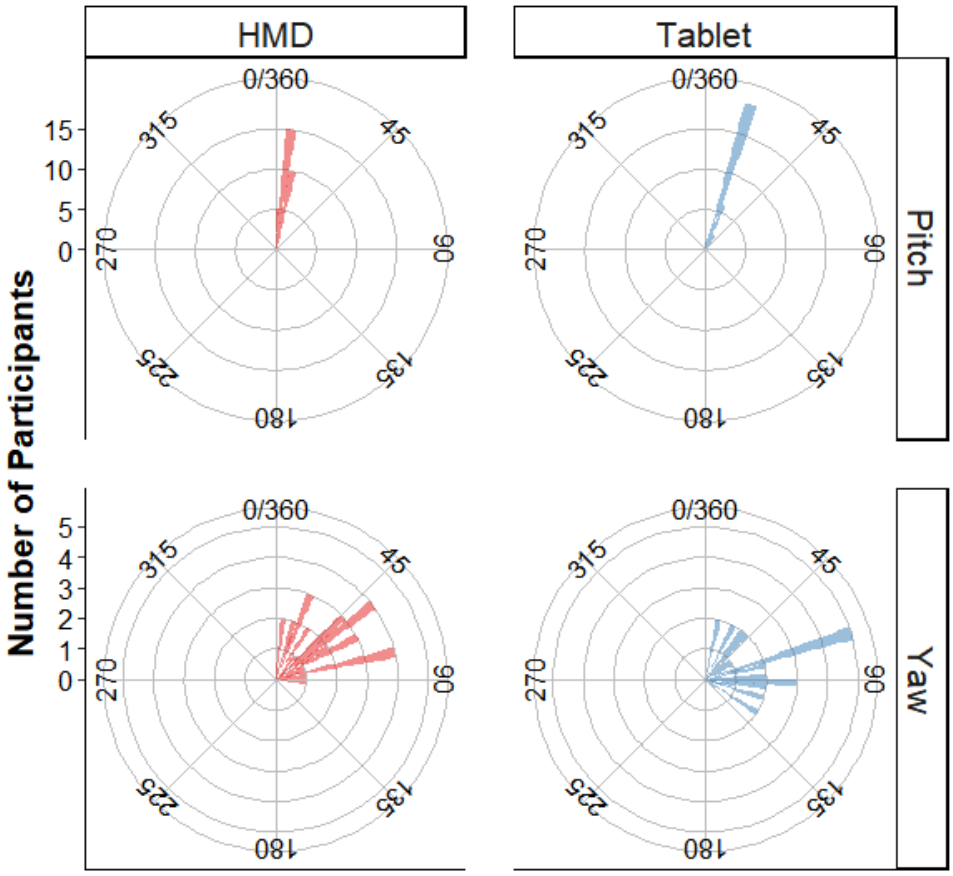

Figure 3. Plots showing the distribution of participants mean viewing direction ($\underline{\theta}$) across the whole viewing session for the pitch (top) and yaw (bottom) axes for the HMD (left) and Tablet (right) groups. $\underline{\theta}$ values were organised in bins every 5°, longer lines from the centre represent a larger number of participants in that bin.

To investigate whether the two device groups differed in their average GCD a Welch two sample t-test comparing the HMD and Tablet groups' mean distance across the session. This revealed no significant difference between the HMD (Mean = 1.28, SD = 0.23) and tablet

(Mean = 1.39, SD = 0.30) groups, t(58.03) = -1.75, p = .086, η2 = 0.08. Levene's test for homogeneity of variance indicated that there was no significant difference in variance between the two groups (F = 0.27, p = .607).

Table 2. Descriptive statistics for viewing angle data with mean direction ($\bar{\theta}$), mean resultant length ($\bar{R}$), circular variance (Vm) and circular standard deviation (v) for each device.

Axis	Device	$\bar{\theta}$	$\bar{R}$	Vm	v
Pitch	HMD	9.22°	.997	.003	1.84°
	Tablet	19.29°	.996	.004	1.47°
Yaw	HMD	51.57°	.919	.081	2.21°
	Tablet	66.48°	.828	.172	2.03°

ISC Analysis

Inter-subject Correlation Across the 360° video

To test whether our groups showed significant ISC, we first compared whether the mean ρ for GCD in each group differed from 0 using one sample t-tests. As Table 3 shows ISC for both devices was significantly greater than 0.

To discover if ISC significantly differed between devices across the whole study, we ran a Welch two sample t-test comparing the ISC of the HMD and Tablet groups. This revealed a significant difference between groups, $t(61.13) = 2.89$, $p = .005$, $\eta^2 = 0.17$, (see Figure 4).

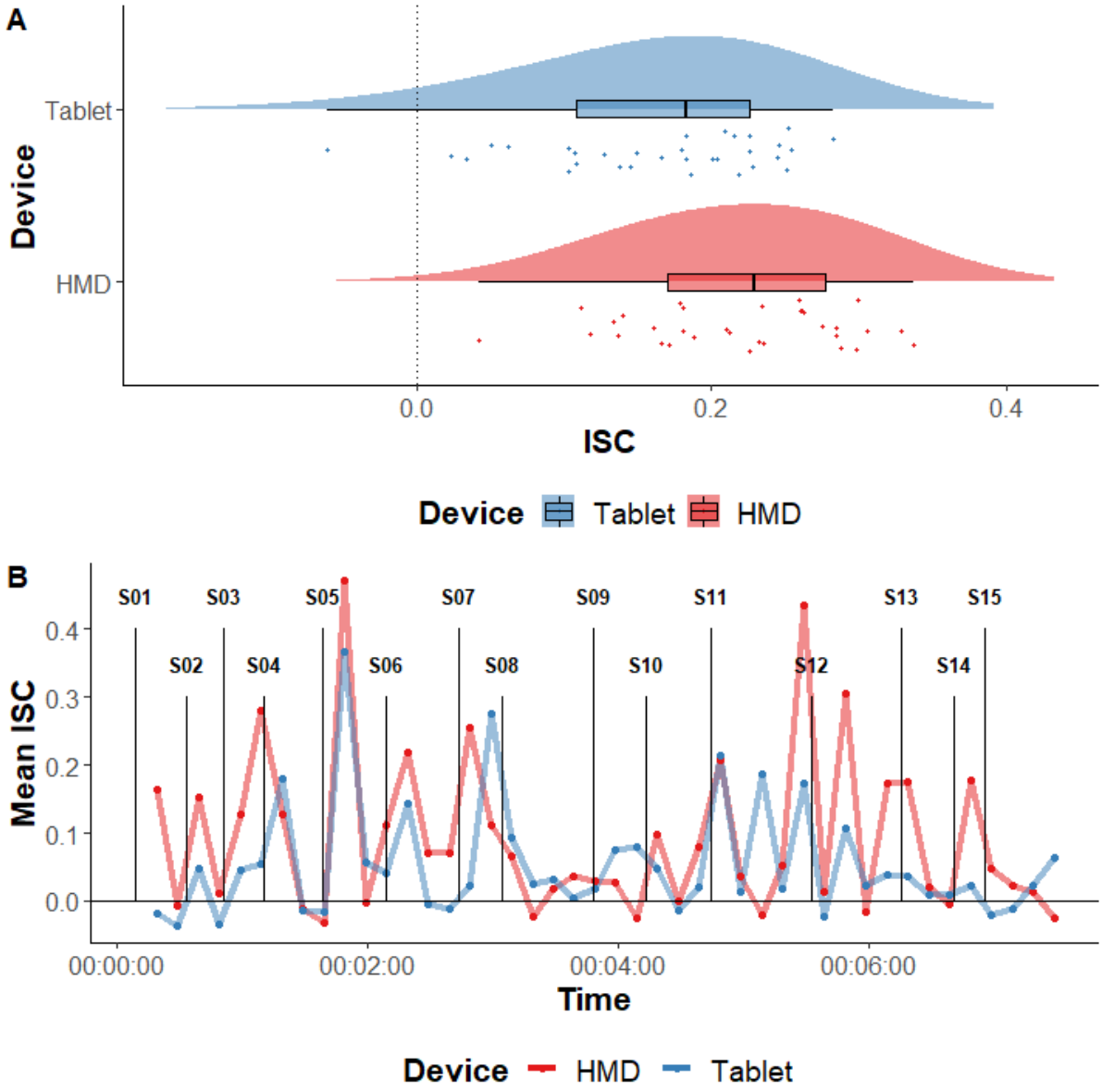

Figure 4. A) Mean ISC across the whole viewing session across device. Clouds represent distribution, raindrops represent individual participants. B) Results of an analysis of ISC in 10 second bins across the entire time period. Points represent mean ISC for the preceding 10 seconds. Markers represent scene starts.

Table 3. Mean ISC and standard deviation (SD) along with the t statistic (t) and degrees of freedom (df) for one sample t-tests comparing against 0 for each group.

Device	Mean ISC	SD	t	df
HMD	0.219	.071	17.34	31
Tablet	0.163	.081	9.88	31

Inter-subject Correlation by Scene Type

To investigate how directional cues within scenes affected ISC, we carried out a 2x3 ANOVAs with device (HMD vs Tablet) as the between subject factor and scene category (Unidirectional vs Intermediate vs Panoramic) as the within subjects factor (see Table 4 and Figure 5).

Table 4. Means and standard deviations of ISC across scene types, device and measure. Same letter superscripts represent a significant difference between devices within scene type and measure in post hoc t-tests of estimated marginal means, † indicates $p < .01$.

Device	Scene Type	Mean ISC	SD
HMD	Unidirectional	.225$^{\dagger a}$	.088
	Intermediate	.152$^{\dagger b}$	.089
	Panoramic	.012	.023
Tablet	Unidirectional	.151$^{\dagger a}$	.012
	Intermediate	.085$^{\dagger b}$	.092
	Panoramic	.030	.051

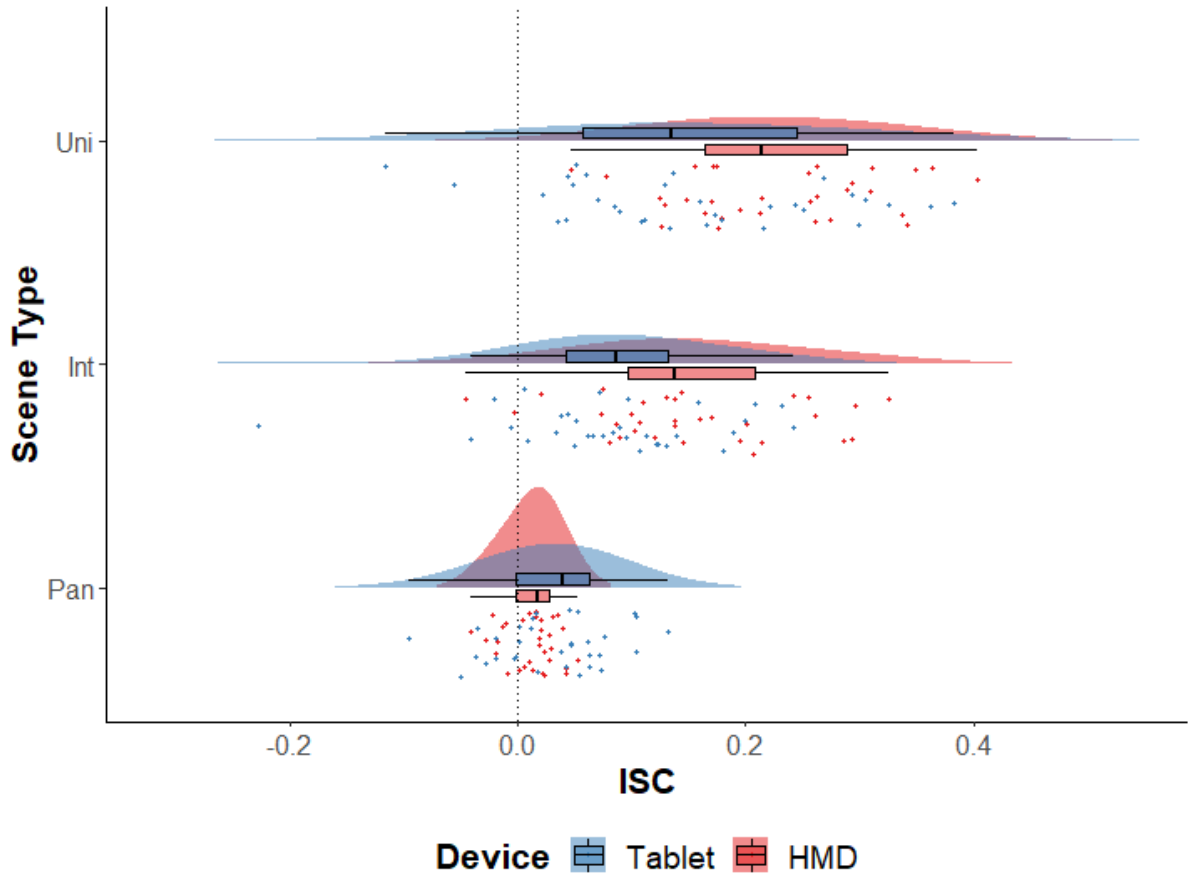

Figure 5. Mean ISC across scene type and device. Clouds represent distribution, raindrops represent individual participants. Uni equals Unidirectional, Int equals intermediate and Pan equals Panoramic.

We found a significant effect of device, $F(1,62) = 9.15$, $p = .004$, $\eta_p^2 = 0.13$, which was driven by higher ISC in the HMD group compared to the Tablet group. There was also a significant effect of scene type, $F(1.82,113.04) = 74.09$, $p < .001$, $\eta_p^2 = 0.54$ which was driven by significantly higher ISC in Unidirectional scenes compared to Intermediate scenes, $t(189) = -4.49$, $p < .001$, and Panoramic scenes, $t(189) = -10.70$, $p < .001$ and significantly higher ISC in Intermediate scenes compared to Panoramic scenes, $t(124) = -6.25$, $p < .001$. There was also a significant interaction between device and scene type, $F(1.62,100.42) = 6.78$, $p = .002$, $\eta_p^2 = 0.10$. Holm corrected post hoc tests revealed that this interaction was driven by significantly greater ISC in Unidirectional scenes, for the HMD group compared to

the Tablet group, $t(186) = 3.53$, $p = .001$, and significantly greater ISC in Intermediate scenes for the HMD group compared to the Tablet group, $t(186) = 3.19$, $p = .002$. No significant difference was found in Panoramic scenes between HMD and Tablet groups, $t(186) = -0.83$, $p = .405$.

Inter-subject Correlation by Time Period Within Scene

To investigate how the time period within scenes affected ISC, we carried out a 2x3 ANOVA with device (HMD vs Tablet) as the between subject factor and time period (Beginning vs Middle vs End) as the within subjects factor (see Table 5 and Figure 6).

Table 5. Means and standard deviations of ISC across time periods, device and measure. Same letter superscripts represent a significant difference between devices within time period and measure in post hoc t-tests of estimated marginal means, ‡ indicates $p < .001$.

Device	Time Period	Mean ISC	SD
HMD	Start	.206‡a	.063
	Middle	.050	.040
	End	.051	.038
Tablet	Start	.101‡a	.081
	Middle	.043	.049
	End	.024	.039

We found a significant effect of device, $F(1,62) = 21.40$, $p < .001$, $\eta_p^2 = 0.26$, which was driven by higher ISC in the HMD group compared to the Tablet group. There was also a significant effect of time point, $F(1.67,103.78) = 132.68$, $p < .001$, $\eta_p^2 = 0.682$, which was

driven by significantly higher ISC at the Start of scenes compared to in the Middle, $t(189) = -9.71$, $p < .001$, or at the End, $t(189) = -10.60$, $p < .001$, of scenes. There was no significant difference between ISC in the Middle and End of scenes, $t(189) = -0.84$, $p = .400$. There was also a significant interaction between device and time point, $F(1.67,103.78) = 21.36$, $p < .001$, $\eta_p^2 = 0.26$. Holm corrected post hoc tests revealed that this interaction was driven by significantly greater ISC at the Start of scenes for the HMD group compared to the Tablet group, $t(186) = 7.75$, $p < .001$. By contrast, in the Middle of scenes there was no significant difference in ISC between HMD and Tablet groups, $t(186) = -0.53$, $p = .594$. The difference in ISC between HMD and Tablet groups at the End of scenes approached significance, $t(186) = 1.93$, $p = .055$.

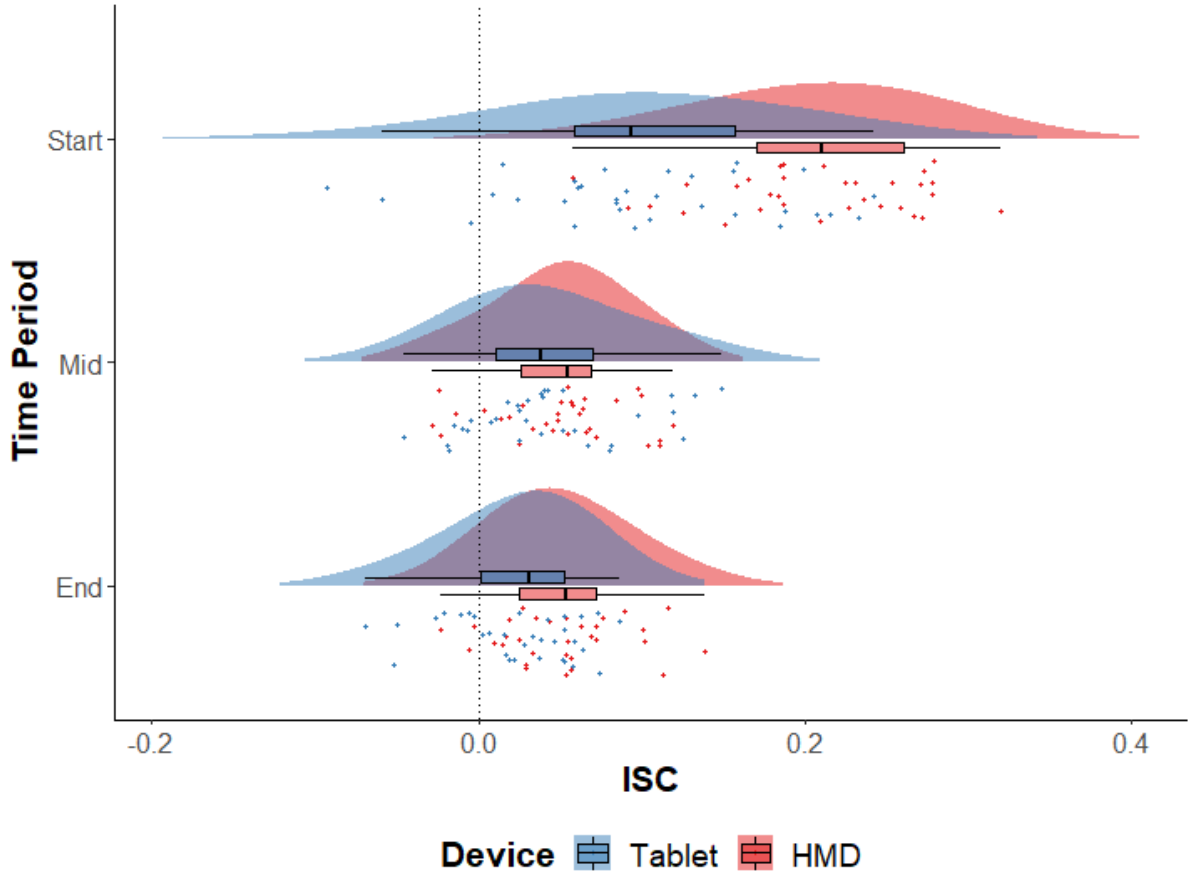

Figure 6. Mean ISC across time point and device. Clouds represent distribution, raindrops represent individual participants.

Relationship between ISC and Engagement with the Documentary

To discover if engagement scores significantly differed between devices, we ran a Welch two sample t-test comparing the HMD and Tablet groups on engagement. This revealed that participants reported significantly higher engagement in the HMD group (Mean = 28.25, SD = 4.67) compared with the tablet group (Mean = 23.47, SD = 7.06), $t(61.13) = 2.89$, $p = .005$, $\eta^2 = 0.17$.

We also investigated the relationship between engagement and ISC by running a regression analysis with engagement as the outcome variable and z-scored GCD as the predictor. A significant regression equation was found, F(1, 62) = 5.54, p = .022, with an R2 of 0.08. Participants' predicted engagement is equal to 25.86 + 1.85 (ISC). Engagement increased 1.85 for each SD increase in ISC.

Discussion

In this section we first discuss our findings and offer possible interpretations before moving on to discuss some of the limitations of the current study and possible directions for future research in this area.

Our analysis of pitch and yaw data revealed significant differences in mean directions ($\underline{\theta}$) for both the pitch and yaw axis with significantly greater distribution of pitch and yaw $\underline{\theta}$ in the Tablet group compared to the HMD group. This finding suggests that participants in the HMD group were more cohesive in their mean viewing direction while those in the Tablet group centred their viewing around a wider variety of points. In the yaw condition, the greater distribution of $\underline{\theta}$ in the Tablet group may be due to the fact that the tablet was wireless - allowing full unencumbered rotational movement in 360° - whereas the HMD required a

cable connection to the testing laptop which may have led users to avoid excessive yaw rotation due to fear of becoming tangled in or pulling out the cable. When we converted pitch and yaw into a single measure of great circle distance however there were no significant differences between the groups in either their mean distance or the variance in mean distance. This suggests that, while participants' mean viewing direction varied more in the Tablet than the HMD group, the average amount of distance that participants' viewpoints moved from the reference point was not significantly different between devices.

The overall inter-subject correlation (ISC) analysis revealed that ISC across both groups was significantly different from 0 indicating that the piece did elicit a level of attentional synchrony in both groups. Further, participants exhibited higher ISC in the HMD as compared to the Tablet condition. However, it should be noted that, while ISC in both conditions were significantly above 0, even the HMD ISC values were considerably lower than those found between adult humans in previous studies that investigated ISC for standard video watching (.219 for HMD in the current study compared to .597 in Franchak et al., 2016; and .39 in Shepherd et al., 2010). This reduction in overall ISC may in part be due to the fact that, while previous studies directly measured gaze from eye tracking, the current study relied on head movement data alone. It may also, however, reflect the additional challenges that 360° video creates for coherent storytelling.

In addition to examining attentional synchrony (AS) over the entire time course we also carried out two additional analyses. The first of these examined whether differences in the directionality of attentional cues in each scene encouraged greater AS. This analysis found that - perhaps unsurprisingly - participants showed the highest level of AS for scenes that had a single direction of focus, with decreased AS in intermediate scenes and the lowest value for panoramic scenes that lacked a clear direction. We also found an interaction between device

and scene, in which participants watching using an HMD showed higher AS than those watching using a Tablet for unidirectional and intermediate scenes but not for panoramic scenes. These findings suggest that HMD users are more responsive to directional cues within a scene than Tablet users, which may help to explain the overall increase in AS for those using HMDs. It is notable that in many of the directional scenes, the item of interest was a person or group of people which previous studies have shown act as strong attentional attractors (Birmingham et al., 2008, 2009; Williams et al., 2018).

As well as investigating how the spatial properties of a scene affected AS we also investigated the influence of temporal properties on AS by comparing the average ISC of 10 second periods at the start, middle and end of scenes. We found significantly greater AS at the beginning of scenes compared to in the middle or at the end. Again, this effect interacted with device type with a stronger effect of time point on AS in our HMD group compared to our Tablet group. The overall pattern of increased AS at the start of scenes is similar to previous findings from 2D film which suggest that AS declines over the course of a scene (Dorr et al., 2010). One explanation for this effect is evidence that, when viewing a new static or dynamic scene, gaze behaviour can be divided into two phases (Eisenberg & Zacks, 2016; Unema et al., 2005). First there is an ambient period, characterised by a reliance on input from peripheral vision and the rapid acquisition of low frequency information. This is followed by a focal period characterised by longer fixation on areas of high frequency information. It is likely that AS is higher during the more feature driven ambient period due to less room for individual differences in interests and motivations to interfere with the capture of visual attention by salient aspects of the scene.

There are several explanations as to why watching 360° videos on a 2D display as opposed to HMD might act to weaken AS. First, unlike HMDs, which fully immerse the viewer within a

visual scene, 2D displays do not fully block out the external environment, which in the current case contained a significant number of different objects and a wide view of part of the Bath University campus. These additional visual features that were present for participants in the Tablet condition may have made them more likely to have been influenced by visual information not within the 360° video (Neumann & Moffitt, 2018) than were participants in the HMD condition.

Second, the full visual immersion offered by HMDs means that more of the scene is seen in peripheral vision compared to a flat screen set up. Since peripheral vision is highly sensitive to motion cues (Strasburger et al., 2011) which have been shown to influence attentional control (Inoue et al., 2017), it is possible that viewers in a HMD will be more likely to respond to peripheral cues than will participants in a flat screen condition. The fact that our effects seemed to be particularly strong at the start of scenes suggests that the greater AS observed in the HMD group during these times is driven by access to saliency cues in peripheral vision that act to guide viewer's attention toward salient features of the scene. On this account, greater access to these saliency cues in the HMD group would lead participants' attention to move in a more synchronous manner during the early exploratory period within each scene (Serrano et al., 2018). Similarly increased peripheral vision might aid participants using an HMD in locating the salient areas within directional scenes.

Our final analysis investigated the relationship between AS and participants' rating of their engagement with the documentary. A linear regression found that ISC was a positive predictor of participants' scores on the engagement questionnaire, suggesting that participants whose viewpoints overlapped more with those of their fellow participants found the documentary more engaging. This finding is important as it indicates that our measures of AS

substantially affect viewers' perceptions of the 360 piece and that increasing the amount of AS found among viewers could increase viewer enjoyment.

In terms of implications our findings suggest several possible strategies for content producers seeking to maximise AS between viewers in VR. First, the finding of stronger AS for unidirectional compared to panoramic scenes suggests the value of avoiding visual "clutter" in non-relevant areas for scenes in which creators intend there to be only one area of interest. Second the findings of stronger AS at the beginning compared to the middle and end of scenes suggest that the content creator could benefit from adding scene transitions just prior to times when they want viewers' attention to shift to a new location in the visual scene.

Limitations

Our study had several limitations in its design. First, we measured attentional direction based on the central point of the participant's view rather than examining actual fixation data using eye-tracking. This means that we cannot be sure that our participants were fixating on the coordinates of the tracking data as opposed to moving their eyes to a different point within the visual scene. However, it should be noted that Sitzmann and colleagues (2018) investigated the amount of coupling between head and eye movements and found evidence of a strong correlation with head movements following eye gaze direction with an average delay of 58ms. They further demonstrated that head movement data alone predicted saliency as well gaze based predictors when viewing 360° scenes. These findings suggest that, at least within an HMD, viewing direction is a suitable proxy for eye-gaze.

A second limitation of our study is that, as mentioned above, there were significant differences in the affordances of the two devices. The presence of the cable attaching the HMD to the laptop may have restrained participant movement in the yaw axis. This problem

could be overcome by use of the new generation of wireless headsets to test AS in future studies. A related issue is that the "magic window" set up of the tablet is not a particularly naturalistic way to watch 360° video on a tablet and may have limited viewers' exploratory behaviour. On the other hand, had we allowed participants to explore the scene using finger movements (as most 360° video viewers for flat screens do) then there would have been even greater differences in visuo-motor affordances between the two devices than in the present setup. We would also note that the analyses showing modulation of AS by both the temporal and spatial properties of scenes were strongest in the HMD condition meaning that these findings have relevance for those working in VR content production independently of the comparison with other non-HMD formats.

A final limitation of this study is that it is not clear the extent to which the different degrees of ISC seen in our results translate into actionable intelligence about the coherence of scenes or the effectiveness of attentional cueing. While this question will require further research looking specifically at ISC in 360 video to answer, we note that the positive relationship we found between ISC and engagement suggests that higher AS can positively affect viewers' perceptions of 360 video. We also note that although our overall levels of ISC were lower than those found in previous studies that used ISC to study AS of gaze direction the differences in ISC found between the HMD and Tablet devices and between scene types and time periods on the HMD conditions were large suggesting that these factors are likely to play a meaningful role in modulating AS.

Future Directions

While our study offers an early step towards understanding AS in 360° video, there are several directions that future research could take. One obvious advance on the current study would be to directly collect gaze data in addition to head/tablet orientation in order to have a

more fine-grained measure of AS. Another potential future direction would be to manipulate the aspects of different scenes or pieces more systematically. For example, would a deliberately structured 360° video lead to stronger AS that an unstructured one? Do we see the same pattern of AS when observing landscape scenes as opposed to scenes populated with people? Another approach would be to use reverse correlation analysis (Hasson et al., 2004) in order to extract the time points that led to the greatest ISC. If this approach was combined with more quantitative measures of scene complexity and annotation it might yield additional insights into the scene features that promote AS in 360° video.

Another question is how much the co-presence of viewers affects AS. Golland and colleagues (Golland et al., 2015) showed that watching a 2D film with another person increased the amount of ISC in physiological measures. While HMDs are a particularly solipsistic medium, the growing interest in VR cinema (Kil, 2018) raises the question of how much co-presence with others acts to synchronise attention. There are also questions as to how AS is related to measures of neural and physiological synchrony and to viewer enjoyment of (and immersion in) the piece, all of which are ripe for further investigation.

Conclusion

The rapid uptake of VR by producers of nonfiction reflects a fascination with the potential of 360 media for viewer engagement. While VR allows viewers a novel freedom to find their own visual pathway through a given scene, at the same time producers have particular information that they need to convey. The power to elicit AS is often held up as a key indicator of a content creator's ability to lead an audience through a narrative. In this paper we have shown that 360° video can elicit a degree of AS between viewers and that this effect is increased when the video is viewed via a HMD. This effect was greater for scenes with a clear direction of focus than for more panoramic scenes and greater at the onset of scenes

than at the middle or end. These insights may aid content producers as to how to ensure that they structure scenes such that viewers are most likely to see crucial features. Finally, we hope that showing the feasibility of this method of analysis encourages other researchers to consider applying a similar analysis to their own 360° video tracking data and help to develop an empirically informed "screen grammar" for virtual reality.

Online Material

The raw datafile and the R scripts used for the analyses reported in this study are available online at https://osf.io/axgb5/.

References

Adolphs, R., Nummenmaa, L., Todorov, A., & Haxby, J. V. (2016). Data-driven approaches in the investigation of social perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693). https://doi.org/10.1098/rstb.2015.0367

Agostinelli, C., & Lund, U. (2017). *R package "circular": Circular Statistics (version 0.4-93)* (0.4-93). https://r-forge.r-project.org/projects/circular/

Bender, S. (2018). Headset attentional synchrony: Tracking the gaze of viewers watching narrative virtual reality. *Media Practice and Education*. https://doi.org/10.1080/25741136.2018.1464743

Bevan, C., & Green, D. P. (2017). A mediography of virtual reality non-fiction: insights and future directions. *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '18)*, 161–166. https://doi.org/10.1145/3210825.3213557

Bevan, C., Green, D. P., Farmer, H., Rose, M., Cater, K., Stanton Fraser, D., & Brown, H. (2019). Behind the curtain of the "ultimate empathy machine": On the composition of virtual reality nonfiction experiences. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 509. https://doi.org/10.1145/3290605.3300736

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, *49*(24), 2992–3000. https://doi.org/10.1016/j.visres.2009.09.014

Birmingham, E., Bischof, W., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition*, *16*(2–3), 341–355. https://doi.org/10.1080/13506280701434532

Bracken, B. K., Alexander, V., Zak, P. J., Romero, V., & Barraza, J. A. (2014). Physiological synchronization is associated with narrative emotionality and subsequent behavioral response. In D. D. Schmorrow & F. C. M (Eds.), *Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems. AC 2014. Lecture Notes in Computer Science, vol 8534. Springer, Cham* (pp. 3–13). Springer. https://doi.org/https://doi.org/10.1007/978-3-319-07527-3_1

Breathnach, D. (2016). Attentional synchrony and the effects of repetitive movie viewing. *CEUR Workshop Proceedings*, *1751*, 260–271.

Burleson-Lesser, K., Morone, F., DeGuzman, P., Parra, L. C., & Makse, H. A. (2017). Collective behaviour in video viewing: A thermodynamic analysis of gaze position. *PLoS ONE*, *12*(1), 1–19. https://doi.org/10.1371/journal.pone.0168995

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, *46*(26), 4333–4345.

https://doi.org/10.1016/j.visres.2006.08.019

consultancy.uk. (2018). *Virtual and augmented reality market to boom to $170 billion by 2022*. https://www.consultancy.uk/news/17876/virtual-and-augmented-reality-market-to-boom-to-170-billion-by-2022

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 1–17. https://doi.org/10.1167/14.8.5

Cremers, J., & Klugkist, I. (2018). One direction? A tutorial for circular data using R with examples in cognitive psychology. *Frontiers in Psychology*, *9*, Article 2040. https://doi.org/10.3389/FPSYG.2018.02040

Dmochowski, J. P., Bezdek, M. A., Abelson, R. P., Johnson, J. S., Schumacher, E. H., & Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications*, *5*, 1–9. https://doi.org/10.1038/ncomms5567

Dooley, K. (2017). Storytelling with virtual reality in 360-degrees: A new screen grammar. *Studies in Australasian Cinema*, *11*(3), 161–171. https://doi.org/10.1080/17503175.2017.1387357

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28–28. https://doi.org/10.1167/10.10.28

Eco VR. (2017). *Virtual Nature 360° - 5K Nature Meditation for Daydream, Oculus, Gear VR*. https://www.youtube.com/watch?v=rG4jSz_2HDY

Eisenberg, M. L., & Zacks, J. M. (2016). Ambient and focal visual processing of naturalistic activity. *Journal of Vision*, *16*(2), 5. https://doi.org/10.1167/16.2.5

1 Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical
2 power analysis program for the social, behavioral, and biomedical sciences. *Behavioural*
3 *Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

4 Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free viewing gaze
5 behavior in infants and adults. *Infancy*, *25*(3), 289–313.
6 https://doi.org/110.1016/j.bbi.2017.04.008

7 Golland, Y., Arzouan, Y., & Levit-Binnun, N. (2015). The mere co-presence:
8 Synchronization of autonomic signals and emotional responses across co-present
9 individuals not engaged in direct interaction. *PLoS ONE*, *10*(5), Article e0125804.
10 https://doi.org/10.1371/journal.pone.0125804

11 Golland, Y., Keissar, K., & Levit-Binnun, N. (2014). Studying the dynamics of autonomic
12 activity during emotional experience. *Psychophysiology*, *51*(11), 1101–1111.
13 https://doi.org/10.1111/psyp.12261

14 Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced intersubject
15 correlations during movie viewing correlate with successful episodic encoding. *Neuron*,
16 *57*(3), 452–462. https://doi.org/10.1016/j.neuron.2007.12.009.Enhanced

17 Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008).
18 Neurocinematics: The neuroscience of film. *Projections*, *2*(1), 1–26.
19 https://doi.org/10:3167/proj.2008.020102 Volume

20 Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject
21 synchronization of cortical activity during natural vision. *Science*, *303*, 1634–1640.
22 https://doi.org/10.1126/science.1089506

Herbec, A., Kauppi, J. P., Jola, C., Tohka, J., & Pollick, F. E. (2015). Differences in fMRI intersubject correlation while viewing unedited and edited videos of dance performance. *Cortex*, *71*, 341–348. https://doi.org/10.1016/j.cortex.2015.06.026

Hijmans, R., Williams, E., & Vennes, C. (2019). *Package ' geosphere '* (1.5-10). https://CRAN.R-project.org/package=geosphere.

HTC. (2016). *Vive*. https://www.vive.com

Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cognitive Research: Principles and Implications*, *2*, Article 46. https://doi.org/10.1186/s41235-017-0080-5

Inoue, Y., Tanizawa, T., Utsumi, A., Susami, K., Kondo, T., & Takahashiy, K. (2017). Visual attention control using peripheral vision stimulation. *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, *2017-Janua*, 1363–1368. https://doi.org/10.1109/SMC.2017.8122803

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, *12*(6), 1093–1123. https://doi.org/10.1080/13506280444000661

Kil, S. (2018). *Virtual Reality Comes to Mainstream Cinema in South Korea*. Variety. https://variety.com/2018/biz/news/vr-virtual-reality-gina-kim-1202801584-1202801584/

Kirkorian, H. L., & Anderson, D. R. (2018). Effect of sequential video shot comprehensibility on attentional synchrony: A comparison of children and adults. *Proceedings of the National Academy of Sciences*, *115*(40), 9867–9874.

https://doi.org/10.1073/pnas.1611606114

Lankinen, K., Saari, J., Hari, R., & Koskinen, M. (2014). Intersubject consistency of cortical MEG signals during movie viewing. *NeuroImage*, *92*, 217–224. https://doi.org/10.1016/j.neuroimage.2014.02.004

Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS ONE*, *10*(11), 1–23. https://doi.org/10.1371/journal.pone.0142474

MacQuarrie, A., & Steed, A. (2017). Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. *2017 IEEE Virtual Reality (VR)*, 45–54. https://doi.org/10.1109/VR.2017.7892230

Margerie, T. de. (2018). *How to integrate 360 video with Unity*. https://blogs.unity3d.com/2017/07/27/how-to-integrate-360-video-with-unity/?_ga=2.159851645.1308616920.1543498780-421605029.1517835635#

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, *3*(1), 5–24. https://doi.org/10.1007/s12559-010-9074-z

Neumann, D., & Moffitt, R. (2018). Affective and attentional states when running in a virtual reality environment. *Sports*, *6*(3), 71. https://doi.org/10.3390/sports6030071

Oculus. (2016). *Oculus Rift*. https://www.oculus.com/rift

Oculus. (2019). *Oculus Quest*. https://www.oculus.com/quest/

Ozcinar, C., & Smolic, A. (2018). Visual attention in omnidirectional video for virtual reality

applications. *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 1–6. https://doi.org/10.1109/QoMEX.2018.8463418

Passmore, P. J., Glancy, M., Philpot, A., Roscoe, A., Wood, A., & Fields, B. (2016). Effects of viewing condition on user experience of panoramic video. In D. Reiners, D. Iwai, & F. Steinicke (Eds.), *Proceedings of the 26th International Conference on Artificial Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments* (pp. 9–16). The Eurographics Association. https://doi.org/10.2312/egve.20161428

Pope, V. C., Dawes, R., Schweiger, F., & Sheikh, A. (2017). The geometry of storytelling: Theatrical use of space for 360-degree videos and virtual reality. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4468–4478. https://doi.org/10.1145/3025453.3025581

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, *7*, 1–9. https://doi.org/10.1038/srep43916

Rubo, M., & Gamer, M. (2018). Social content and emotional valence modulate gaze fixations in dynamic scenes. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-22127-w

Samsung. (2015). *Gear VR*. https://www.samsung.com/global/galaxy/gear-vr/

Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion*, *41*(6), 708–712. https://doi.org/10.1007/s11031-017-9641-7

Serrano, A., Sitzmann, V., Ruiz-Borau, J., Wetzstein, G., Gutierrez, D., & Masia, B. (2018). Movie editing and cognitive event segmentation in virtual reality video. *ACM*

Transactions on Graphics, *36*(4), 47. https://doi.org/10.1145/3072959.3073668

Shepherd, S. V, Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human–monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Current Biology*, *20*(7), 649–656. https://doi.org/10.1016/j.cub.2010.02.032.Human

Silver, N. C., & Dunlap, W. P. (1987). Averaging Correlation Coefficients: Should Fisher's z Transformation Be Used? *Journal of Applied Psychology*, *72*(1), 146–148. https://doi.org/10.1037/0021-9010.72.1.146

Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., & Wetzstein, G. (2018). Saliency in VR: How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), 1633–1642. https://doi.org/10.1109/TVCG.2018.2793599

Smith, T. J. (2013). Watching you watch movies: Using eye tracking to inform cognitive film theory. In A. P. Shimamura (Ed.), *Psychocinematics* (pp. 165–192). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199862139.003.0009

Smith, T. J., Levin, D., & Cutting, J. E. (2012). A window on reality: Perceiving edited moving images. *Current Directions in Psychological Science*, *21*(2), 107–113. https://doi.org/10.1177/0963721412437407

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*(8), 1–24. https://doi.org/10.1167/13.8.16

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13–13. https://doi.org/10.1167/11.5.13

Subramanian, R., Shankar, D., Sebe, N., & Melcher, D. (2014). Emotion modulates eye movement patterns and subsequent memory for the gist and details of movie scenes. *Journal of Vision*, *14*(3), 31–31. https://doi.org/10.1167/14.3.31

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494. https://doi.org/10.1080/13506280444000409

VRSE.works. (2015). *Clouds over Sidra*. https://www.youtube.com/watch?v=mUosdCQsMkM&t=3s

Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, *12*(1), 16–16. https://doi.org/10.1167/12.1.16

Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior*, *32*, 123–132. https://doi.org/10.1016/j.chb.2013.12.001

Wikipedia. (2020). *Great-circle distance*. Wikipedia. https://en.wikipedia.org/wiki/Great-circle_distance

Williams, E. H., Cristino, F., & Cross, E. S. (2018). Human body motion captures visual attention and elicits pupillary arousal. *Cognition*, *193*(November), 104029. https://doi.org/10.1016/j.cognition.2019.104029

Zoric, G., Barkhuus, L., Engström, A., & Önnevall, E. (2013). Panoramic video: Design challenges and implications for content interaction. *Proceedings of the 11th European*

Conference on Interactive TV and Video (EuroITV '13), 153–162. https://doi.org/10.1145/2465958.2465959